

На правах рукописи

СМИРНОВ ЮРИЙ ВИКТОРОВИЧ

**ИНФОРМАЦИОННЫЙ ПОИСК ДЛЯ ОБЛАЧНЫХ БИБЛИОТЕЧНЫХ
СИСТЕМ: ОСОБЕННОСТИ ЛИНГВИСТИЧЕСКОГО ОБЕСПЕЧЕНИЯ**

Специальность 05.25.05 – Информационные системы и процессы

АВТОРЕФЕРАТ

диссертации на соискание учёной степени
кандидата технических наук

Москва
2020

Диссертация выполнена в Федеральном государственном бюджетном образовательном учреждении высшего образования «Московский государственный институт культуры», на кафедре библиотечно-информационных наук.

Научный руководитель: **Шрайберг Яков Леонидович,**
доктор технических наук, профессор, научный руководитель ФГБУ
«Государственная публичная научно-техническая библиотека
России» (ГПНТБ России),
г. Москва

**Официальные
оппоненты:** **Залаев Геннадий Захарович,**
доктор технических наук, профессор, начальник отдела
Автоматизированных архивных технологий
ФКУ «Российский государственный архив научно-технической
документации» (РГАНТД),
г. Москва

Лаврёнова Ольга Александровна,
кандидат филологических наук, начальник отдела методического и
лингвистического обеспечения информационных систем
ФГБУ «Российская государственная библиотека» (ФГБУ «РГБ»),
г. Москва

Ведущая организация: Федеральное государственное бюджетное учреждение науки
Всероссийский институт научной и технической информации
Российской академии наук (ВИНИТИ РАН)

Защита диссертации состоится «27» марта 2020 г. в 14 часов 00 минут на заседании совета по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук Д 210.010.01, созданного на базе Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный институт культуры», по адресу: 141406, Московская обл., г. Химки-6, ул. Библиотечная, д. 7, корп. 2, зал защиты диссертаций (218 ауд.).

С диссертацией можно ознакомиться в Информационно-библиотечном центре и на сайте Московского государственного института культуры: <http://nauka.mgik.org/>

Автореферат разослан «___» _____ 2020 г.

Ученый секретарь диссертационного совета,
кандидат педагогических наук, доцент



Т. Я. Кузнецова

I. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Повсеместное распространение интернета дало начало появлению облачных технологий, внедрение которых на сегодняшний день происходит во многих сферах деятельности: в медицине, в бизнесе, в государственном управлении, в образовании и т. д. Библиотеки также начали внедрение этих технологий, позволяющих информации преодолевать географические и технологические барьеры. Для получения доступа к информации о фондах библиотек пользователям можно физически не присутствовать в библиотеке, достаточно иметь компьютер и доступ в интернет.

На базе облачных технологий начали создаваться **облачные библиотечные системы (ОБС)** – тип автоматизированных библиотечно-информационных систем, использующий облачные технологии и распространяемый по модели «Программное обеспечение как услуга» (SaaS).

Одной из важнейших причин медленного внедрения облачных технологий в библиотеках является несоответствие программного обеспечения нуждам библиотеки, в частности – отсутствие удовлетворительного тематического поиска. Облачные библиотечные системы (ОБС), бесспорно, имеют встроенные поисковые возможности, но это либо полнотекстовый поиск, либо поиск по небольшому количеству полей базы данных (например: поиск по автору, заглавию, году издания и т. д.).

Главной задачей поиска, как известно, является удовлетворение информационной потребности пользователя, для чего используется различное лингвистическое обеспечение (например: словарные базы данных, информационно-поисковые языки, в том числе классификационные УДК, ББК и др.).

Эффективность и скорость информационного поиска зависит от качества индексирования документов, т. е. от описания документов средствами информационно-поискового языка (ИПЯ).

Использование таких ИПЯ в ОБС часто является для пользователей затруднительным, поскольку эти языки обладают довольно большим словарем (например: Универсальная десятичная классификация содержит 68194 индексов¹ или словарных единиц), который усложняет сам процесса поиска, поскольку вынуждает пользователя осуществлять дополнительный поиск одной или нескольких тем по этому словарю.

¹ Гиляревский, Р. С. Рубрикатор как инструмент информационной навигации / Р. С. Гиляревский, А. В. Шапкин, В. Н. Белоозеров. – Санкт-Петербург : Профессия, 2008. – С. 269.

Также стоит отметить, что на разработку и последующее внедрение классификационных ИПЯ влияет необходимость заключения и последующего продления платных лицензионных договоров с правообладателями на таблицы классификаций.

Несмотря на то, что не все библиотеки спешат перейти с привычных АБИС на ОБС, часть из них постепенно переходит, что оказывает влияние и на растущие требования к этим системам, которые пока уступают по функциональности традиционным. Одной из недоработок ОБС является недостаточная для библиотек разработанность лингвистического обеспечения информационного поиска, на которую влияют следующие факторы:

- 1) большое разнообразие типов ИПЯ, процедуры поиска которых необходимо реализовывать по-разному;
- 2) отсутствие стандартизованного набора ИПЯ для всех библиотек страны;
- 3) необходимость платного лицензирования для использования некоторых ИПЯ.

Поскольку модель ОБС предполагает интенсивную работу с огромными массивами как структурированных, так и неструктурированных данных, которые сегодня можно отнести к классу так называемых «Больших данных» (Big Data), для лингвистического обеспечения открываются новые возможности. Одной из таких возможностей является интеграция всех видов ИПЯ, поскольку чем больше различных средств поиска предоставляется пользователю, тем лучше будет удовлетворена его информационная потребность.

Исходя из вышесказанного, можно с уверенностью утверждать, что разработка интегративной модели тематического поиска для облачных библиотечных систем является актуальной.

Степень разработанности темы исследования. Информационно-поисковыми языками, их типологией и использованием в информационно-поисковых системах занимались такие ученые, как В. И. Москович, А. И. Черный, М. А. Довбенко, А. Н. Уманский, А. В. Соколов, В. П. Захаров, Э. Р. Сукиасян, В. Н. Белоозеров, О. А. Лавренова, Е. М. Зайцева и другие. Советский и российский опыт разработки лингвистического обеспечения в электронных библиотеках был обобщен А. Б. Антопольским.

Развитию теоретических основ также способствовали диссертации Н. И. Гендиной, И. Л. Скипор, Г. А. Скарук, Л. Н. Пирумовой, Г. В. Кульковой, А. Ш. Меркулова. Однако, в последнее время интерес к этой теме заметно снизился, что подтверждает небольшое количество публикаций по данной теме. Например, за последние 18 лет была защищена всего одна диссертация, посвященная тематическому поиску.

Наиболее известными учеными в области совместимости ИПЯ являются Г. В. Гейнц, создавший комплексно-системный каталог, Р. С. Гиляревский, В. Н. Белоозеров и др.

Вопросы, связанные с организацией поисковых систем, были затронуты в работах Я. Л. Шрайберга, Ф. С. Воройского, А. Б. Антопольского, Н. Е. Каленова и др.

Исследования в области облачных технологий начали появляться с 2008 года, а в России – с 2010 года, однако ни одно из них не было посвящено их применению в библиотечном деле. Однако, это не означает, что этим вопросам не уделяется внимание. Об использовании облачных технологий в библиотеках пишут Рупеш Шанчати, И.В. Билан, Р. С. Гиляревский, Е. В. Ковязина, В. В. Трофимова и др.

Значительный вклад в развитие проблематики облачных систем внес Я. Л. Шрайберг, который одним из первых ученых обратил внимание на необходимость использования облачных технологий в библиотеках. В своих ежегодных докладах на международных конференциях и форумах «Крым», начиная с 2009 года, он продолжает уделять этому внимание.

Проведенный анализ показывает, что существуют многочисленные работы, с различных сторон освещающие проблематику лингвистического обеспечения информационного поиска, тематического поиска и облачных технологий. При этом прямая постановка вопроса о разработке модели тематического поиска для облачных библиотечных систем в литературе отсутствует.

Целью настоящего исследования является совершенствование библиотечных процессов информационного поиска с помощью современных облачных технологий.

Задачи исследования. Достижение поставленной цели требует решения следующих задач:

- проанализировать существующие средства организации информационного поиска;
- определить современное состояние лингвистического обеспечения информационного поиска в облачных библиотечных системах;
- проанализировать использование облачных технологий в библиотечных системах и определить понятие «облачная библиотечная система»;
- теоретически обосновать и построить интегративную модель тематического поиска для облачных библиотечных систем, сформулировать основные требования к модели тематического поиска для облачных библиотечных систем;
- разработать лингвистическое обеспечение интегративной модели тематического поиска для облачных библиотечных систем.

Научная новизна определяется оригинальностью и теоретической обоснованностью решений в рамках актуальной задачи современной информационной науки и практики –

необходимостью внедрения облачных библиотечных систем как важнейшего компонента современной информационной инфраструктуры:

- введено и определено понятие: «облачная библиотечная система»;
- предложена научная концепция информационного обеспечения тематического поиска для облачных библиотечных систем;
- разработана интегративная модель тематического поиска для облачных библиотечных систем, отличающаяся от существующих интеграцией классификационного и дескрипторного подходов при организации тематического поиска;
- обосновано использование УДК в качестве классификационного ИПЯ в интегративной модели тематического поиска, а также произведено сокращение таблиц УДК эвристическим подходом с использованием определенной логики
- доказана перспективность использования разработанной интегративной модели для организации тематического поиска в облачных библиотечных системах.

Полученные новые результаты являются значительными улучшениями ранее существовавших, новые информационно-технологические решения позволяют устранить причины замедления темпов внедрения облачных технологий в библиотечную практику.

Теоретическая значимость работы заключается в том, что:

- доказаны положения, расширяющие представление информационных наук о возможностях реализации облачных технологий в библиотечной практике, раскрыты проблемы, замедляющие их эффективное применение в библиотечном деле, предложены обладающие научной новизной решения этих проблем;
- уточнен понятийный аппарат информационных наук – дано научно-обоснованное определение понятия «облачные библиотечные системы»;
- проведена модернизация существующих подходов к организации информационно-поисковых систем электронных библиотек;
- разработана, теоретически обоснована и результативно апробирована модель тематического поиска, интегрирующая классификационный и дескрипторный подходы и позволяющая оптимизировать поисковые процессы и повысить качество информационного поиска в автоматизированных информационно-библиотечных системах.

Практическая значимость диссертации заключается:

- в предложенном автором решении практических проблем, связанных с эффективным функционированием облачных технологий в практике информационно-библиотечной

деятельности, которое содействует реализации отдельных позиций национальной программы «Цифровая экономика» в части федерального проекта «Информационная инфраструктура».

- в научно-обоснованном определении понятия «облачная библиотечная система», которое может использоваться для подготовки учебной и методической литературы, а также государственных стандартов Системы стандартов по информации, библиотечному и издательскому делу (СИБИД);
- в разработке системы практических рекомендаций для сокращения количества индексов классификационного ИПЯ (на примере УДК) эвристическим подходом;
- в возможности практического применения разработанной автором интегративной модели тематического поиска в облачных библиотечных системах, что позволит решить проблему недостаточной разработанности информационного обеспечения тематического поиска.

В процессе работы над диссертацией автором была разработана «Электронный терминологический словарь по СИБИД. А-Я», а также опубликованы методические рекомендации «Создание небиблиографических баз данных для обеспечения технологических потребностей библиотек», которые применяются библиотеками страны и используются в образовательном процессе Московского государственного института культуры при реализации основной профессиональной образовательной программы по направлению подготовки 51.03.06 - Библиотечно-информационная деятельность.

Методология и методы исследования.

Результаты проведенных исследований, научных положений и выводов диссертации основываются на фундаментальных положениях современной науки и техники, отраженных в приведенных обзорах научной литературы по проблемам, связанным с информационным поиском, используемым лингвистическим обеспечением и облачными технологиями, и подтверждаются экспериментальными и теоретическими исследованиями, проведенными автором. Для решения поставленных задач исследования применялись следующие методы исследования: классификационный анализ, сравнительный анализ, метод моделирования.

Объектом исследования является информационный поиск в облачной библиотечной системе.

Предметом исследования является лингвистическое обеспечение информационного поиска для облачных библиотечных систем.

Положения, выносимые на защиту. На защиту диссертации выносятся следующие ключевые положения:

1) научное обоснование понятия «облачная библиотечная система – это тип автоматизированных библиотечно-информационных систем, использующий облачные технологии и распространяемый по модели «Программное обеспечение как услуга» (SaaS)»;

2) основными требованиями, предъявляемыми к интегративной модели тематического поиска для облачных библиотечных систем, являются: независимость от естественного языка, снижение влияния недостатков естественного языка на поисковый запрос, простота использования, наглядность структуры, группировка по любому сочетанию признаков, управление полнотой выдачи, простота обновления и модернизации, возможность увеличения емкости ИПЯ, гибкость структуры, легкость обучения, удобство осуществления поиска;

3) разработанная интегративная модель тематического поиска представляет собой сочетание классификационного подхода для выражения основной темы документа и дескрипторного для уточнения тематики документа.

Соответствие диссертации паспорту научной специальности. Тема и содержание диссертации соответствуют научной специальности 05.25.05 - Информационные системы и процессы и технической отрасли науки согласно пункту 5 паспорта научной специальности 05.25.05 - Информационные системы и процессы в части «Организационное обеспечение информационных систем и процессов, в том числе новые принципы разработки и организации функционирования информационных систем и процессов... Общие принципы и основы организации информационных служб и электронных библиотек. Стандартизация информационного и лингвистического обеспечения».

Степень достоверности и обоснованность результатов обеспечены применением широкого комплекса методов, адекватных решаемой научной задаче, в том числе методов детального анализа отечественных и зарубежных автоматизированных библиотечно-информационных систем по вопросам изучаемой проблемы, практической реализацией результатов исследования, обширным количеством научных публикаций автора диссертации.

Апробация результатов исследования. По теме диссертации сделано 23 публикации, из них 5 – в рецензируемых научных журналах, рекомендованных Высшей аттестационной комиссией при Министерстве науки и высшего образования Российской Федерации для опубликования основных научных результатов диссертаций на соискание учёных степеней доктора и кандидата наук; индексированных в международных реферативных базах данных – 5 (Web of Science -4, Scopus -1).

Отдельные положения и результаты исследования изложены на российских и международных научных и научно-практических конференциях и форумах:

- 1) Международная конференция «Информационные технологии, компьютерные системы и издательская продукция для библиотек» («ЛИБКОВ») (г. Суздаль, 2015 - 2018 гг.);
- 2) Международная научно-практическая конференция «Румянцевские чтения» (г. Москва, 2016 г., 2017 г., 2019 г.);
- 3) «Скворцовские чтения» — XXI Международная научная конференция «Библиотечное дело — 2016: библиотечно-информационные коммуникации в поликультурном пространстве» (г. Москва, 27 - 28 апреля 2016 г.);
- 4) Международный профессиональный форум "Книга. Культура. Образование. Инновации" (г. Судак, Республика Крым, 2016 - 2019 гг.);
- 5) Международная конференция к 65-летию ВИНТИ РАН «Информация в современном мире» (г. Москва, 25 - 26 октября 2017 г.);
- 6) Научно-практическая конференция «Перспективные направления научных исследований и критические технологии в классификационных системах» (г. Москва, 25 - 26 октября 2017 г.);
- 7) «Скворцовские чтения» — XXII Международная научная конференция «Библиотечное дело — 2019. Библиотека в цифровой среде: тенденции развития» (г. Москва, 19 апреля 2019 г.)

Экспериментальной базой исследования послужил электронный каталог авторефератов диссертаций ФГБУ «Государственной публичной научно-технической библиотеки России» (ГПНТБ России).

Структура диссертации. Цели и задачи диссертации обусловили ее структуру. Работа состоит из введения, трех глав, заключения, списка использованных сокращений, списка литературы (131 наименование) и 2 приложений. В тексте работы 16 рисунков и 5 таблиц. Общий объем диссертации составляет 228 страниц.

II. ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснована актуальность темы исследования, выяснена степень ее разработанности, определены объект, предмет, цель и задачи, теоретическая и практическая значимость исследования, раскрыта научная новизна, сформулированы положения, выносимые на защиту.

Первая глава «Современное состояние лингвистического обеспечения информационного поиска для облачных библиотечных систем» состоит из трех разделов. **В первом разделе** проведен анализ основных средств организации информационного поиска: каталоги, которые являются основным поисковым средством в библиотеках, и поисковые системы интернета.

В контексте данного исследования особое внимание было уделено методикам индексирования электронных каталогов и поисковых систем интернета, а также следующим системам: уточнения запросов, автоматического перевода, предметизации и каталогизации. В ходе анализа было установлено, что методики индексирования в поисковых системах интернета и библиотеках отличаются. Первые создают ПОД автоматически, с помощью слов и словосочетаний, используемых в документе, и некоторого набора метаданных (например, дата документа), вторые – с помощью специальных информационно-поисковых языков (ИПЯ).

Второй раздел посвящен современному состоянию лингвистического обеспечения информационного поиска. В этом разделе рассмотрены различные информационно-поисковые языки (ИПЯ): вербальные, лексика которых представляет собой слова и выражения на естественном языке пользователя (например: язык ключевых слов, тегирование), и кодированные, словарь которых представлен кодами (например: УДК, ББК, ГРНТИ).

Проведенный анализ лингвистического обеспечения показал, что при организации информационного поиска активно используется лингвистическое обеспечение от небольших словарей до информационно-поисковых языков, однако существенных различий в использовании определенных ИПЯ в поисковых системах интернета и библиотеках нет. Например, полнотекстовый поиск используется как в поисковых системах, так и в библиотеках. Однако существуют некоторые предпочтения используемого лингвистического обеспечения: если для организации информационного поиска библиотеки используют традиционные средства – различные типы ИПЯ, как в карточных, так и электронных каталогах, то поисковые системы интернета предпочитают использовать различные системы перевода с одного естественного языка на другой, а для систематизации информации – тегирование.

Третий раздел посвящен облачным технологиям и их использованию в библиотеках, в нем раскрывается понятие облачных технологий, виды моделей по типу развертывания (частное облако, публичное, общественное и гибридное) и по типу обслуживания (SaaS — программное обеспечение как услуга, PaaS — платформа как услуга, IaaS — инфраструктура как услуга), а также их основные характеристики (самообслуживание по требованию, универсальный доступ по сети, объединение ресурсов, эластичность, учет потребления).

Рассмотрены виды автоматизированных библиотечно-информационных систем и определено понятие **облачная библиотечная система (ОБС)** – это тип автоматизированных библиотечно-информационных систем, использующий облачные технологии и распространяемый по модели «Программное обеспечение как услуга» (SaaS).

Уделено особое внимание влиянию проводимой правительством РФ политики импортозамещения программного обеспечения, согласно которой в государственных и муниципальных предприятиях России (в том числе, библиотеках) можно использовать только российское ПО. Этим требованиям отвечают следующие ОБС: «МАРК Cloud», АИБС «МегаПро» и САБ ИРБИС64/128.

В ходе исследования установлено, что эти ОБС (кроме САБ ИРБИС64/128) предлагают несколько видов поиска (например: по основным полям, расширенный и полнотекстовый поиск и т. д.), однако возможности систематического поиска, являющегося очень важным для библиотек, ограничены, что нивелирует все преимущества облачных вычислений.

Проведенный анализ средств информационного поиска, используемого лингвистического обеспечения и возможностей ОБС при организации тематического поиска приводит к заключению, что разработка интегративной модели тематического поиска является не только возможной, но и необходимой для решения поставленных в исследовании задач.

Вторая глава «Разработка модели лингвистического обеспечения информационного поиска для облачных библиотечных систем» состоит из двух разделов. **В первом разделе** обосновываются основные требования к разрабатываемой для ОБС интегративной модели тематического поиска, для чего потребовалось установить причины необходимости этой разработки:

- 1) отсутствие тематического поиска в ОБС, характерного для многих АБИС, поскольку на сегодняшний день он представлен в ОБС в виде посимвольного поиска по полям базы данных (БД);
- 2) необходимость заключения и продления лицензионных договоров с правообладателями большинства ИПЯ.

Для формулировки требований к модели интеграции были исследованы:

- 1) модели тематического поиска (дескрипторная, предметная, фасетная и иерархическая), их достоинства и недостатки;
- 2) ИПЯ (язык ключевых слов, тегирование, алфавитно-предметная классификация, УДК, ББК и ГРНТИ), которые можно использовать в интегративной модели, их достоинства и недостатки;
- 3) основные принципы организации пользовательского интерфейса тематического поиска, используемые на практике (строка поиска, выпадающий список, выбор поля базы данных для поиска, словарь терминов, иерархическое дерево), принципы отображения информации в поисковом интерфейсе: скрытие ненужной пользователю информации (разработка БЕН РАН для УДК, при которой структура УДК выглядит как алфавитно-предметная классификация) и вербальное представление иерархического поиска (разработка РГБ, которая заключается в использовании языка ключевых слов, основанном на словесном выражении индексов ББК, и

реализуется с помощью средств семантической паутины (Semantic Web)), их достоинства и недостатки.

В результате анализа различных моделей тематического поиска, достоинств и недостатков ИПЯ, в также принципов организации пользовательского интерфейса, автором был составлен список основных требований, предъявляемых к интегративной модели:

- 1) независимость от естественного языка, которой обладают кодированные ИПЯ, – для поиска документов на разных языках;
- 2) снижение влияния недостатков естественного языка на поисковый запрос, как это организовано в кодированных ИПЯ;
- 3) простота использования, которой обладают вербальные ИПЯ, за счет использования слов естественного языка при составлении поискового запроса;
- 4) наглядность структуры, которой обладают иерархические ИПЯ, – для предоставления возможностей навигации по ИПЯ;
- 5) группировка по любому сочетанию признаков, характерная для фасетных и фасетно-иерархических ИПЯ;
- 6) управление полнотой выдачи, как у иерархических ИПЯ, посредством расширения или сужения поискового запроса;
- 7) простота обновления и модернизации, характерная для вербальных ИПЯ – для предоставления актуального состояния отраслей науки;
- 8) возможность увеличения емкости ИПЯ;
- 9) гибкость структуры, как у фасетных ИПЯ;
- 10) легкость обучения, вплоть до интуитивного понимания «что делать?» и «как искать?», для привлечения большего числа пользователей применять на практике предлагаемую поисковую систему;
- 11) удобство осуществления поиска, которое зависит от пользовательского интерфейса.

Проведенное исследование показало, что найти «золотую середину» между потребностями пользователя и возможностями ИПЯ и поисковых систем возможно, если объединить преимущества различных типов ИПЯ, с одновременным снижением количества их недостатков.

Второй раздел посвящен разработке интегративной модели тематического поиска.

Описаны основные причины неудачного внедрения новых ИПЯ, которые могли бы устранить недостатки уже существующих языков или моделей тематического поиска:

- 1) недостаточно полная разработанность словаря ИПЯ;
- 2) труднореализуемая на практике система поиска посредством классификационного ИПЯ;

- 3) отсутствие поддержки со стороны государства или библиотечного сообщества.

Уделено особое внимание лицензированию ИПЯ, и описаны условия использования UDC Summary под открытой лицензией Creative Commons Attribution Share Alike 3.0:

- 1) разрешается свободно копировать и распространять материал;
- 2) разрешается создавать производные материалы на основе предоставляемого, в том числе для коммерческих целей;
- 3) требуется обязательно указать авторство, ссылку на лицензию и обозначить изменения;
- 4) производный материал обязательно должен распространяться на условиях этой же лицензии;
- 5) нельзя применять юридические или технические меры защиты производного материала, противоречащие условиям лицензии.

В ходе анализа систем лицензирования было установлено, что использование UDC Summary позволяет воспользоваться уже готовой иерархической структурой классификационного ИПЯ, не привлекая к этой работе специалистов, чей труд необходимо было бы оплатить. Также ее большим достоинством является отсутствие необходимости в заключении лицензионного договора и в его дальнейшем продлении.

Исследование словаря UDC Summary показало, что он содержит меньше лексических единиц по сравнению с полными таблицами УДК, однако отражает все отрасли и крупнейшие направления науки. Автор приходит к выводу, что для более подробной детализации можно воспользоваться одним из вербальных языков, что, конечно, может приводить к ошибкам и дублирующим понятиям, которые, однако, не повлияют на основную тематику, выраженную с помощью индекса УДК из UDC Summary.

Основываясь на требованиях, описанных в предыдущем разделе, и предлагаемых решениях основных проблем внедрения, автором выявлены основные принципы построения интегративной модели тематического поиска:

- 1) основой этой модели должен быть кодированный ИПЯ для обеспечения: независимости результатов поиска от естественного языка пользователя; наглядности структуры ИПЯ; возможности управлять полнотой выдачи, т. е. сужать и расширять поисковый запрос;
- 2) словарь лексических единиц интегративного ИПЯ должен быть основан на уже существующем кодированном ИПЯ для обеспечения: снижения затрат на составление словаря и разработку структуры; совместимости с используемыми в библиотеках классификациями;
- 3) словарь лексических единиц этого ИПЯ должен быть представлен только основными индексами классификации, которые отражают основные направления науки;

4) более дробная детализация должна производиться с использованием вербального ИПЯ (язык ключевых слов, тегирование) с нерегулируемой лексикой для обеспечения: простоты использования и поиска необходимого понятия; легкости обучения этому ИПЯ; простоты обновления и модернизации; возможности увеличения емкости ИПЯ;

5) группировка по сочетанию признаков может производиться как с использованием вспомогательных таблиц кодированного ИПЯ (УДК, ББК) или различных рубрикаторов (например: для обозначения страны можно использовать Общероссийский классификатор стран мира), так и с помощью подрубрик, выраженных ключевыми словами, что обеспечит гибкость структуры ИПЯ.

Таким образом, в этой модели сочетаются все основные виды моделей тематического поиска: дескрипторная, предметная, фасетная и иерархическая.

В общем виде интегративную модель можно представить следующим образом: основное направление науки, к которому тематически относится документ, отображается индексом уже существующей и используемой в России классификации (УДК, ББК или ГРНТИ), а более детально тема документа выражается с помощью ключевых слов.

Для выявления наиболее подходящей классификации в качестве основы интегративной модели были проанализированы используемые в настоящее время в России классификации (УДК, ББК, ГРНТИ) по следующим критериям:

- 1) многоязычность, то есть наличие соответствий для нескольких языков;
- 2) наличие синонимов для рубрик;
- 3) логичность структуры (все ли нижестоящие индексы входят в вышестоящие);
- 4) практический опыт использования в библиотеках России, т. е. наличие компетентных специалистов;
- 5) сочетаемость, т. е. возможность внедрения в структуру исходного ИПЯ других видов ИПЯ;
- 6) лицензионная открытость, т. е. отсутствие необходимости заключать лицензионные соглашения с правообладателями.

По результатам проведенного анализа (см. таблицу 1) делается вывод, что УДК более всего соответствует всем критериям и является наиболее подходящим вариантом ИПЯ для создания на ее основе интегративной модели тематического поиска, однако это не означает, что ББК и ГРНТИ не могут быть использованы для этих целей.

Таблица 1. Таблица соответствия ГРНТИ, ББК и УДК заявленным критериям

	ГРНТИ	ББК	УДК
Многоязычность	Нет (только русский)	Нет (только русский)	Да (57 языков)

Наличие синонимов для рубрик	Есть для немногих рубрик	Есть для многих рубрик	Есть, но для немногих рубрик
Логичность структуры	Да	Да	Да
Практический опыт использования	Да	Да	Да
Сочетаемость	Нет	Нет необходимости	Есть, с помощью знака «*» (Звездочка)
Лицензионная открытость	Нет	Нет	Полные таблицы – Нет. UDC Summary – Да.

Несмотря на то, что UDC Summary содержит примерно 2600 индексов, для простоты использования интегративной модели их количество было сокращено до наименований наук и наиболее важных понятий с использованием эвристического подхода, а именно:

- 1) из таблиц на русском языке были исключены индексы и пояснения на английском, поскольку они затрудняют поиск индекса по ключевым словам в рубрике;
- 2) были исключены из таблиц индексы, не являющиеся наименованиями наук и наиболее важными понятиями;
- 3) текст исключенных рубрик, содержащий важные при поиске ключевые слова, был перемещен в расширения рубрик вышестоящих индексов для улучшения поиска по синонимам;
- 4) были удалены ссылки на исключенные рубрики;
- 5) количество вспомогательных таблиц было сокращено до таблиц наименований стран, языков, народов и видов документов.

После проведенной автором корректировки количество индексов УДК в таблицах для интегративной модели сократилось более чем в два раза: в UDC Summary было 2600 индексов, а в откорректированной версии таблиц стало 1100 индексов: около 650 индексов основных таблиц, содержащих основные научные понятия, и около 450 индексов вспомогательных таблиц, включающих наименования стран, языков, народов и видов документов.

Таким образом, основная тема документа будет выражена с помощью индекса УДК из сокращенной таблицы, который является одним из основных направлений науки. Для детализации темы документа предлагается применять дескрипторный ИПЯ (язык ключевых слов или тегирование) с использованием некоторых принципов построения предметных рубрик для нормализации формулировок ключевых слов.

Исходя из вышесказанного, были сформулированы основные правила детализации ключевыми словами для интегративной модели тематического поиска:

- 1) ключевое слово начинается символом «#» (знак решетки);

- 2) слова в словосочетании разделяются пробелом;
- 3) два ключевых слова разделяются символом «#» (знак решетки);
- 4) слово или словосочетание должно быть в словарной форме, а именно: в именительном падеже единственного или множественного числа;
- 5) выраженное ключевым словом понятие должно быть зафиксировано в документах;
- 6) ключевое слово должно формулироваться с учетом потребностей и удобства пользователей;
- 7) каждое следующее ключевое слово, по необходимости, должно ограничивать объем понятия предыдущего.

Исходя из этих правил, были сформулированы основные этапы работы с интегративной моделью тематического поиска и ее принципиальная схема. Особенностью интегративной модели тематического поиска является интеграция (объединение) классификационного и дескрипторного подходов при организации тематического поиска, что проявляется как в процессе индексирования, так и в процессе поиска. Предлагаемая модель моделирует два процесса: процесс индексирования, который отражает описание документов в терминах информационно-поискового языка, и процесс поиска, который отражает поиск документов, соответствующих поисковому запросу, введенному пользователем.

Схема процесса индексирования состоит из 6 этапов:

- 1) анализ содержания документа;
- 2) выделение основной темы документа;
- 3) поиск основной темы документа в словаре УДК, который может осуществляться двумя способами: навигацией по иерархическому дереву или поиском индексов в словаре классификатора по ключевым словам рубрик;
- 4) занесение найденных индексов в соответствующее поле базы данных ОБС, определяемое правилами каталогизирующей организации, разделяя их знаком «#» (решетка): либо в подполя поля, предназначенного для индекса УДК, либо в поле, предназначенное для индексов другой классификации;
- 5) выделение основных ключевых слов;
- 6) занесение ключевых слов в соответствующее поле базы данных ОБС, определяемое правилами каталогизирующей организации, разделяя их знаком «#» (решетка) либо в поле, предназначенное для ключевых слов, либо в поле, предназначенное для индексов другой классификации, добавляя к проставленному ранее индексу УДК и разделяя все элементы разделителем «#» (знак решетки).

Поскольку в большинстве баз данных АБИС присутствуют поля для различных ИПЯ, используемых в библиотеках, то интегративная модель тематического поиска не требует внесения изменений в структуру БД.

Схема процесса поиска состоит из следующих шагов:

- 1) формулировка поискового запроса производится пользователем по таблицам классификации и по ключевым словам в списке ключевых слов;
- 2) разделение поискового запроса на элементы;
- 3) проверка каждого элемента поискового запроса:
 - 3.1) если элемент является индексом УДК, то он добавляется в массив индексов УДК поискового запроса, с помощью которого в дальнейшем будет производиться поиск в соответствующем поле БД, определяемом правилами каталогизирующей организации;
 - 3.2) если элемент является ключевым словом, то он добавляется в массив ключевых слов поискового запроса, с помощью которого в дальнейшем будет производиться поиск в соответствующем поле БД, определяемом правилами каталогизирующей организации;
- 4) в соответствующих полях каждого документа БД, определяемых правилами каталогизирующей организации, осуществляется проверка на соответствие каждого элемента из массивов ключевых слов и индексов УДК поискового запроса, сформированных ранее, поисковому образу документа с учетом булевой логики «И» или «ИЛИ»;
- 5) независимо от количества найденных документов производятся следующие операции для получения списка ссылок, которая будет отражена в области уточнения результатов поиска:
 - 5.1) исследование массива индексов УДК поискового запроса:
 - 5.1.1) при наличии вышестоящего индекса у каждого индекса УДК из поискового запроса он добавляется в раздел «Расширить область поиска»;
 - 5.1.2) при наличии нижестоящих индексов у каждого индекса УДК из поискового запроса они добавляются в раздел «Сузить область поиска»;
 - 5.1.3) при наличии сопоставленных с каждым индексом УДК из поискового запроса используемых ключевых слов они добавляются в раздел «Ключевые слова, используемые с индексами УДК»;
 - 5.2) исследование массива ключевых слов поискового запроса:
 - 5.2.1) при наличии сопоставленного индекса УДК у каждого ключевого слова из поискового запроса этот индекс добавляется в раздел «Уточнить тему ключевых слов».
- 6) оценка результатов поиска, которая зависит от количества документов в поисковой выдаче:
 - 6.1) если результат поиска не выявил или выявил слишком мало документов, то поиск может быть расширен двумя способами:

- 6.1.1) переходом вверх по иерархическому дереву, т. е. к вышестоящему индексу, с последующим возвратом к оценке результатов поиска (см. п. 2);
- 6.1.2) формулированием новой, более широкой, темы запроса, с последующим переходом к поиску темы в таблицах классификации (см. п. 1);
- 6.2) если результат поиска выявил слишком много документов, то необходимо сузить область поиска с помощью ключевых слов, используемых для детализации документов с выбранным индексом классификации, после чего вернуться к оценке результатов поиска (см. п. 2);
- 6.3) если результат поиска удовлетворяет пользователя, то процесс поиска может быть завершен.

Выявлены основные достоинства предлагаемой интегративной модели тематического поиска (наглядность словаря ИПЯ, совместимость с традиционными ИПЯ, простота использования, простота обновления и актуализации, соблюдение лицензионной чистоты, упрощение структуры УДК для пользователя) и недостатки (неполная независимость от естественного языка пользователя, неполная совместимость с УДК, определенная трудность в реализации поискового интерфейса).

Третья глава «Экспериментальная оценка эффективности использования интегративной модели тематического поиска» состоит из двух разделов. **Первый раздел** посвящен построению экспериментальной базы для оценки эффективности использования интегративной модели тематического поиска, для проведения которой был использован массив данных из ГПНТБ России и создано программное обеспечение на основе разработанной автором принципиальной схемы интегративной модели тематического поиска. ПО реализует различные возможности поиска, в том числе стандартный поиск по УДК (или другой классификации), осуществляемый по аналогии с другими АБИС и поиск с помощью интегративной модели тематического поиска.

Необходимо отметить, что рядом с результатами поиска с использованием интегративной модели тематического поиска располагается панель навигации по поисковому запросу для уточнения результатов поиска, в которой отображаются как индексы УДК, так и ключевые слова, связанные с поисковым запросом не только иерархически (вышестоящие и нижестоящие индексы), но и содержательно.

Во втором разделе производилась оценка эффективности использования интегративной модели тематического поиска, исходя из результатов которой, можно сделать следующие выводы:

- 1) Точность стандартного поиска по УДК и поиска с помощью интегративной модели тематического поиска почти во всех случаях неизменна.
- 2) Полнота поиска при использовании только ключевых слов как при стандартном поиске, так и при поиске с использованием интегративной модели остается неизменной.

3) Полнота поиска при использовании УДК как отдельно, так и совместно с ключевыми словами колеблется при стандартном поиске, но остается неизменной при использовании интегративной модели тематического поиска. Это объясняется тем, что часть документов индексируется составными индексами, т.е. объединением нескольких индексов с помощью соединительных знаков (например: «:» – знак простого соотношения), что при стандартном поиске игнорируется. Поиск таких документов можно осуществить только с помощью расширенного поиска, указав условие соответствия «содержит», что, однако, может привести к появлению нерелевантных поисковому запросу результатов. В интегративной модели тематического поиска такие документы не игнорируются, поскольку при автоматическом преобразовании присвоенного индекса УДК в индекс УДК из сокращенной таблицы производится разделение составных индексов на простые.

Также следует отметить, что поиск с использованием интегративной модели тематического поиска осуществляется в одной поисковой строке и не требует от пользователя переключать вид поиска с УДК на ключевые слова и обратно. В ОБС стандартные виды поиска по УДК и ключевым словам осуществляются отдельно и требуют от пользователя переключения вида поиска. Одновременный поиск по УДК и ключевым словам в ОБС осуществляется с помощью расширенного поиска, в котором отсутствуют вспомогательные элементы в виде словарей, что требует от пользователя профессиональных навыков поиска.

III. ЗАКЛЮЧЕНИЕ

В заключении подводятся итоги проделанной работы, намечаются основные тенденции и направления дальнейшего изучения данной проблемы.

В результате проведенного диссертационного исследования:

1) введено и определено понятие: «облачная библиотечная система», уточняющее понятийный аппарат теории организации электронных библиотек, которое может использоваться для подготовки учебной и методической литературы, а также Системы стандартов по информации и издательскому делу (СИБИД);

2) предложена научная концепция информационного обеспечения тематического поиска для облачных библиотечных систем, которая в перспективе может использоваться для улучшения информационного обеспечения тематического поиска в облачных библиотечных системах;

3) определены основные требования, предъявляемые к разработке интегративной модели тематического поиска;

4) разработана интегративная модель тематического поиска для облачных библиотечных систем, особенностью которой является интеграция (объединение) классификационного и дескрипторного подходов при организации тематического поиска, что проявляется как в процессе индексирования, так и в процессе поиска;

5) обосновано использование УДК в качестве классификационного ИПЯ в интегративной модели тематического поиска, а также произведено сокращение таблиц УДК эвристическим подходом с использованием определенной логики;

6) разработана система практических рекомендаций для сокращения количества индексов классификационного ИПЯ (на примере УДК) эвристическим подходом;

7) доказана перспективность использования разработанной интегративной модели для организации тематического поиска в облачных библиотечных системах.

Направления дальнейших разработок должны предусматривать исследования по автоматическому индексированию поступающих в библиотеку документов.

Все поставленные задачи и цель диссертационного исследования – совершенствование библиотечных процессов информационного поиска с помощью современных облачных технологий – достигнута.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в изданиях, рекомендованных ВАК России:

1. Смирнов, Ю. В. Облачные вычисления: история и влияние на будущее библиотек / Ю. В. Смирнов // Научные и технические библиотеки. – 2016. – №6. – С. 62-73. – DOI: 10.33186/1027-3689-2016-6-62-73. – WOS:000407884800004.

2. Смирнов, Ю. В. Поисковые системы интернета и методы повышения качества обработки запросов при поиске научной информации в сети / Ю. В. Смирнов // Научные и технические библиотеки. – 2016. – №9. – С. 81-89. – DOI: 10.33186/1027-3689-2016-9-3-54. – WOS:000407885200005.

3. Рожнов, В. И. Переход на российское программное обеспечение в библиотеках / В. И. Рожнов, Ю. В. Смирнов // Научные и технические библиотеки. – 2018. – № 2. – С. 26-36. – DOI: 10.33186/1027-3689-2018-2-26-36. – WOS:000431102500003. Доля вклада автора диссертации – 50%.

4. Смирнов, Ю. В. Использование классификационных информационно-поисковых языков в облачных библиотечно-информационных системах / Ю. В. Смирнов // Научно-техническая информация. Сер.1. Организация и методика информационной работы. – 2018. – № 2. – С. 25-29.

Публикации в изданиях, индексируемых в международных библиографических базах:

1. Smirnov, Yu. V. The use of classification information - retrieval languages in cloud library Information systems / Yu. V. Smirnov // Scientific and Technical Information Processing. – 2018. – Vol.45, №1. – p. 35-39. – DOI: 10.3103/s0147688218010070. – WOS:000433131500005.

Публикации в других изданиях:

1. Зайцева, Е. М. Структурные проблемы Универсальной десятичной классификации (на примере раздела «Российская Федерация. Россия») [Электронный ресурс] / Е.М. Зайцева, Ю.В. Смирнов // Информационные технологии, компьютерные системы и издательская продукция для библиотек («LIBCOM-2015»): материалы конф. – Электрон. текстовые дан. – Режим доступа: <http://gpntb.ru/libcom15/tezis/013.pdf>. Доля вклада автора диссертации - 50%.

2. Смирнов, Ю.В. Электронные версии классификационных систем: их виды и функциональность / Ю.В. Смирнов // Румянцевские чтения – 2016: материалы междунар. науч.-практ. конф. Рос. гос. б-ки, 12-13 апреля 2016 г.: в 2 ч. – Москва: Пашков дом, 2016. – Ч. 2. – С. 140-144.

3. Смирнов, Ю.В. Научный поиск в интернете [Электронный ресурс] / Ю.В. Смирнов // Библиотечное дело – 2016: Библиотечно-информационные коммуникации в поликультурном пространстве. Скворцовские чтения: материалы XXI междунар. конф., 27-28 апреля 2016 г., Москва / Моск. гос. институт культуры – М., 2016. – С. 591-594. – 1 электрон. опт. диск (CD-ROM).

4. Смирнов, Ю.В. Научный поиск в интернете [Электронный ресурс] / Ю.В. Смирнов // Культура: теория и практика: электрон. науч. журн. – 2016. – № 3 (12) (май – июнь). – Режим доступа: <http://theoryofculture.ru/issues/63/876/>.

5. Смирнов, Ю.В. Сравнительный анализ электронных версий классификационных систем [Электронный ресурс] / Ю.В. Смирнов // Книга. Культура. Образование. Инновации («Крым-2016»): труды Четвертого международного профессионального форума, 4-12 июня 2016 г., Судак, Республика Крым, Россия / Государственная публичная научно-техническая библиотека России. Режим доступа: <http://www.gpntb.ru/win/inter-events/crimea2016/disk/1077.pdf>.

6. Смирнов, Ю.В. Электронный терминологический словарь по СИБИБД [Электронный ресурс] / Ю.В. Смирнов // «Информационные технологии, компьютерные системы и издательская продукция для библиотек» («LIBCOM-2016»): материалы конф. Режим доступа: <http://www.gpntb.ru/libcom16/tezis/005.pdf>.

7. Смирнов, Ю.В. Система стандартов по информации, библиотечному и издательскому делу. Состав. Терминология. Словари / Ю.В. Смирнов // Румянцевские чтения – 2017: 500-летие издания первой славянской Библии Франциска Скорины: становление и развитие культуры книгопечатания:

материалы междунар. науч.-практ. конф. – М.: Пашков дом, 2017. – Ч. 2. – С. 132-135. – ISBN 978-5-7510-0716-4.

8. Электронный терминологический словарь по СИБИБД. А-Я [Электронный ресурс] / Государственная публичная науч.-техническая б-ка России (Москва); сост. Ю.В. Смирнов; ред. Е. М. Зайцева; Гос. публ. науч.-техн. б-ка России. Москва: ГПНТБ России, 2017. – 1 электрон. опт. диск (CD-ROM). – ISBN 978-5-85638-201-2.

9. Смирнов, Ю.В. Особенности разработки электронных терминологических словарей (на примере терминологического словаря по СИБИБД) [Электронный ресурс] / Ю.В. Смирнов // Материалы Третьего Международного профессионального форума «Книга. Культура. Образование. Инновации» («Крым – 2017»), 3-11 июня 2017 г., Судак, Республика Крым, Россия / Гос. публ. науч.-техн. б-ка России. Режим доступа: <http://www.gpntb.ru/win/inter-events/crimea2017/disk/055.pdf>.

10. Смирнов, Ю.В. Средства организации систематического поиска в облачных библиотечно-информационных системах [Электронный ресурс] / Ю.В. Смирнов // Информация в современном мире. Международная конференция посвящается 65-летию ВИНТИ РАН: материалы конф., 25-26 окт. 2017 г., ВИНТИ, г. Москва / О.А. Антошкова – отв. за вып.; ВИНТИ РАН. – М., 2017. – С. 267-272.

11. Смирнов, Ю.В. Систематический поиск в облачных библиотечно-информационных системах / Ю. В. Смирнов // Перспективные направления научных исследований и критические технологии в классификационных системах: науч.-практ. конф., 25-27 окт. 2017 г.: материалы конф./ Всерос. ин-т науч. и техн. информации Рос. акад. наук (ВИНТИ РАН) ; [науч. ред. : Е. Ю. Дмитриева]. – Москва: РФФИ, 2017. – С. 59-64. Режим доступа: <http://www.udcc.ru/MATERIALS/2017/CONFERENCE2017.pdf>. – ISBN 978-5-94577-072-0.

12. Зайцева, Е. М. Современные издания терминологических словарей по Системе стандартов по информации, библиотечному и издательскому делу: печатные и электронные версии [Электронный ресурс] / Е. М. Зайцева, Ю. В. Смирнов // «Информационные технологии, компьютерные системы и издательская продукция для библиотек» : материалы Двадцать первой Международной конференции и выставка «LIBCOM-2017», Владимирская область, г. Суздаль, 20 – 24 ноября 2017 г. / Государственная публичная научно-техническая библиотека России . – 2017. – Электрон. дан. – Режим доступа: <http://www.gpntb.ru/libcom2017/doc1/008.pdf>. Доля вклада автора диссертации – 50%.

13. Смирнов, Ю. В. Особенности разработки базы данных терминологического словаря по СИБИБД для системы автоматизации библиотек ИРБИС64 [Электронный ресурс] / Ю. В. Смирнов //

Книга. Культура. Образование. Инновации («Крым – 2018»): труды Четвертого Международного профессионального форума, 16-24 июня 2018 г., Судак, Республика Крым, Россия / Государственная публичная научно-техническая библиотека России. – Электрон. дан. – Режим доступа: <http://www.gpntb.ru/win/inter-events/crimea2018/disk/044.pdf>.

14. Смирнов, Ю.В. Создание небиблиографических баз данных для обеспечения технологических потребностей библиотек: методические рекомендации / Ю.В. Смирнов; ред. Е. М. Зайцева. – Москва: ГПНТБ России, 2019. – 77 с. : ил. – Предм. указ.: с. 67. – 50 экз. – ISBN 978-5-85638-211-1.

15. Смирнов, Ю.В. Анализ поисковых возможностей Единого реестра российских программ для электронных вычислительных машин и баз данных [Электронный ресурс] / Ю.В. Смирнов // «Информационные технологии, компьютерные системы и издательская продукция для библиотек» («LIBCOM-2018»): материалы конф. – Электрон. текстовые дан. – Режим доступа: <http://www.gpntb.ru/libcom2018/doc1/005.pdf>.

16. Зайцева, Е.М. Сравнительный анализ терминологических словарей по СИБИБ / Е.М. Зайцева, Ю.В. Смирнов // Румянцевские чтения – 2019 – The Rumyantsev readings – 2019: материалы Международной научно-практической конференции (23-24 апреля 2019 г.): [в 3 ч.] / Министерство культуры Российской Федерации, Российская государственная библиотека, Библиотечная ассамблея Евразии. – Москва: Пашков дом, 2019. – Ч. 1. – С. 266-272. Доля вклада автора диссертации – 50%.

17. Смирнов, Ю.В. Импортзамещение в российских библиотеках / Ю.В. Смирнов, В.И. Рожнов // Культура: теория и практика. – 2019. – № 3 (30). – Режим доступа: <http://theoryofculture.ru/issues/106/1244/>. Доля вклада автора диссертации – 50%.

18. Смирнов, Ю. В. Электронные издания по стандартизации, выпускаемые ГПНТБ России [Электронный ресурс] / Ю. В. Смирнов // Книга. Культура. Образование. Инновации («Крым – 2019»): труды Пятого Международного профессионального форума, 8-16 июня 2019 г., Судак, Республика Крым, Россия / Государственная публичная научно-техническая библиотека России – Режим доступа: <http://www.gpntb.ru/win/inter-events/crimea2019/disk/030.pdf>.